Suboptimality of Nonlocal Means
for Images with Sharp Edges

Arian Maleki[1,*], Manjari Narayan[2], Richard G. Baraniuk[3,**]

Dept. of Computer and Electrical Engineering, Rice University, MS-380
6100 Main Street, Houston, TX 77005, USA

Abstract

We conduct an asymptotic risk analysis of the nonlocal means image denoising algorithm for the Horizon class of images that are piecewise constant with a sharp edge discontinuity. We prove that the mean square risk of an optimally tuned nonlocal means algorithm decays according to $n^{-1} \log^{1/2+\epsilon} n$, for an n-pixel image with $\epsilon > 0$. This decay rate is an improvement over some of the predecessors of this algorithm, including the linear convolution filter, median filter, and the SUSAN filter, each of which provides a rate of only $n^{-2/3}$. It is also within a logarithmic factor from optimally tuned wavelet thresholding. However, it is still substantially lower than the the optimal minimax rate of $n^{-4/3}$.

Keywords: Denoising, minimax risk, Horizon class, nonlocal means, linear filter, SUSAN filter, wavelet thresholding

---

   *Corresponding author
   **Principal corresponding author
   Email addresses: `arian.maleki@rice.edu` (Arian Maleki), `manjari@rice.edu` (Manjari Narayan), `richb@rice.edu` (Richard G. Baraniuk)
   URL: `http://www.ece.rice.edu/~mam15/` (Arian Maleki), `http://dsp.rice.edu/dspmember/12` (Manjari Narayan), `http://web.ece.rice.edu/richb/` (Richard G. Baraniuk)
   [1]Phone: +1 713.348.3579; Fax: +1 713.348.5685
   [2]Phone: +1 713.348.2371; Fax: +1 713.348.5685
   [3]Phone: +1 713.348.5132; Fax: +1 713.348.5685

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **24 NOV 2011**	2. REPORT TYPE	3. DATES COVERED **00-00-2011 to 00-00-2011**

4. TITLE AND SUBTITLE **Suboptimality of Nonlocal Means for Images with Sharp Edges**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Rice University,Department of Electrical and Computer Engineering,Houston,TX,77005**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We conduct an asymptotic risk analysis of the nonlocal means image denoising algorithm for the Horizon class of images that are piecewise constant with a sharp edge discontinuity. We prove that the mean square risk of an optimally tuned nonlocal means algorithm decays according to n&#56256;&#56320;1 log1=2+ n, for an n-pixel image with > 0. This decay rate is an improvement over some of the predecessors of this algorithm, including the linear convolution lter median lter, and the SUSAN lter, each of which provides a rate of only n&#56256;&#56320;2=3. It is also within a logarithmic factor from optimally tuned wavelet thresholding. However, it is still substantially lower than the the optimal minimax rate of n&#56256;&#56320;4=3.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**33**	

1. Introduction

1.1. Image denoising

The long history of image denoising is testimony to its central importance in image processing. A wide range of algorithms have been developed, ranging from simple linear convolution and median filtering to total variation denoising [1] and sparsity exploiting algorithms such as wavelet shrinkage [2]. Due to the sensitivity of human visual system to edges, the ability to preserve sharp edges is an important criterion for noise removal algorithms. Therefore Korostelev and Tsybakov proposed a framework to characterize the performance of image denoisers on edges [3]. Based on this framework, we aim to characterize the performance of several denoising algorithms that represent the current state of the art image enhancement techniques. In particular, we will focus on the popular and powerful nonlocal means (NLM) algorithm.

1.2. The minimax framework

In this paper, we are interested in estimating a function $f : [0,1]^2 \to \mathbb{R}$ from noisy pixel level observations. Define $\text{Pixel}(i,j) = [\frac{i}{n}, \frac{i+1}{n}) \times [\frac{j}{n}, \frac{j+1}{n})$, and let $x_{i,j} = \text{Ave}(f \mid \text{Pixel}(i,j))$ be the pixel level averages of f. We observe the samples

$$y_{i,j} = x_{i,j} + z_{i,j},$$

where, $z_{i,j}$ is iid $N(0, \sigma^2)$. The goal is to recover the original pixel values $x_{i,j}$ from the observations $y_{i,j}$, based on some information about the function f. For a given function f and an estimator $\hat{f}$ we define the *risk function* as

$$R_n(f, \hat{f}) = \mathbb{E}\left(\frac{1}{n^2} \sum_i \sum_j (x_{i,j} - \hat{f}_{i,j})^2 \right). \tag{1}$$

The risk can also be written as

$$R_n(f, \hat{f}) = \left(\frac{1}{n^2} \sum_i \sum_j (x_{i,j} - \mathbb{E}\hat{f}_{i,j})^2 \right) + \mathbb{E}\left(\frac{1}{n^2} \sum_i \sum_j (\hat{f}_{i,j} - \mathbb{E}\hat{f}_{i,j})^2 \right), \tag{2}$$

where the first and second terms correspond to the *bias* and *variance* of the estimator $\hat{f}$, respectively.

Let f belong to a class of functions $\mathcal{F}$, e.g., a class of edge-like images that represent edges with different shapes and orientations. The risk defined

in (1) depends on the specific choice of f. We define the risk of an estimator $\hat{f}$ on the class $\mathcal{F}$ as the risk of the worst-case signal, i.e.,

$$R_n(\mathcal{F}, \hat{f}) = \sup_{f \in \mathcal{F}} R_n(f, \hat{f}).$$

The minimax risk over functions in $\mathcal{F}$ is then defined as the risk of the best possible estimator, i.e.,

$$R_n^*(F) = \inf_{\hat{f}} \sup_{f \in F} R_n(f, \hat{f}).$$

The minimax risk is a *lower bound* for the performance of all measurable estimators for signals in $\mathcal{F}$.

In this paper we are interested in the asymptotic setting where the number of pixels $n \to \infty$. For all of the estimators we consider, $R_n(\mathcal{F}, \hat{f}) \to 0$ as $n \to \infty$. Therefore, we consider the *decay rate* of the risk as the performance measure. We will derive the minimax risk for several popular image denoising techniques below.

We will use the following asymptotic notation in this paper.

Definition 1. $f(n) = O(g(n))$ *as* $n \to \infty$, *if and only if there exist* n_0 *and* c *such that for any* $n > n_0$, $|f(n)| \leq c|g(n)|$. *Likewise,* $f(n) = \Omega(g(n))$ *as* $n \to \infty$, *if and only if there exist* n_0 *and* c *such that for any* $n > n_0$, $|f(n)| \geq c|g(n)|$. *Finally,* $f(n) = \Theta(g(n))$, *if* $f(n) = O(g(n))$ *and* $f(n) = \Omega(g(n))$. *We may interchangeably use* $f(n) \asymp g(n)$ *for* $f(n) = \Theta(g(n))$.

Definition 2. $f(n) = o(g(n))$ *if and only if* $\lim_{n \to \infty} \frac{f(n)}{g(n)} = 0$.

1.3. Horizon edge model

Several different image edge models have been developed in the image processing and denoising literature. Here we will use the *Horizon model* that contains piecewise constant images with edges that are smooth in the direction of the edge contour but discontinuous in the direction orthogonal to the edge contour [3, 4]. Specifically, let $H\ddot{o}lder^\alpha(C)$ be the class of Hölder smooth functions on $\mathbb{R}$, defined as follows: $h \in H\ddot{o}lder^\alpha(C)$ if and only if

$$|h^{(k)}(t_1) - h^{(k)}(t_1')| \leq C|t_1 - t_1'|^{\alpha - k},$$

where $k = \lfloor \alpha \rfloor$. Given a one-dimensional smooth edge contour function h, we define the image $f_h : [0,1]^2 \to \mathbb{R}$ as $f_h(t_1, t_2) = \mathbf{1}_{\{t_2 < h(t_1)\}}$, where $\mathbf{1}_{\{\cdot\}}$ is
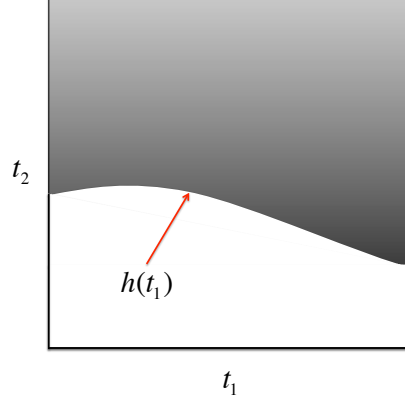
Figure 1: An example of a Horizon function, a piecewise constant image containing an edge that is $Hölder^\alpha$ smooth in the direction of the edge contour but discontinuous in the direction orthogonal to the edge contour.

the indicator function. Based on this construction, we define the Horizon class of functions as

$$H^\alpha(C) = \{f_h(t_1, t_2) : h \in Hölder^\alpha(C) \cap Hölder^1(1)\}, \qquad (3)$$

where α is the smoothness of the edge contour. Figure 1 plots a representative function from this class.

The following theorem, proved in [3], specifies the minimax risk of the class of all measurable estimators on $H^\alpha(C)$.

Theorem 1. [3] *For $\alpha \geq 1$, the minimax risk of the class $H^\alpha(C)$ is*

$$R_n^*(H^\alpha(C)) \asymp n^{\frac{-2\alpha}{\alpha+1}}. \qquad (4)$$

We are particularly interested in the case $\alpha = 2$ edges, for which the optimal rate is $n^{-4/3}$. The rate provided in the above theorem is the Holy Grail of image denoising algorithms.

1.4. A menagerie of denoising algorithms

We will perform a minimax risk analysis of not just nonlocal means but a number of other popular image denoising algorithms.

4

1.4.1. Linear filtering

The classical denoising method is the *linear convolution filter*, which estimates the image via

$$\hat{f}_g^{LF}(i,j) = \sum_m \sum_\ell g_{m,\ell} y_{i-m,j-\ell}, \tag{5}$$

where g is a two dimensional filter impulse response that satisfies $\sum_i \sum_j g_{i,j} = 1$.[4] When all the weights g_{ij} are equal, the algorithm is called the *running average* or the *box filter*. Most of the linear filters used in practice are symmetrical and approximately isotropic.

Definition 3. *Let g be a real and symmetric filter response, i.e., $g_{i,j} = g_{-i,j} = g_{i,-j}$, and let $\hat{G}(\omega_1, \omega_2)$ represent its two-dimensional Fourier transform. The filter is* isotropic *if and only if there exists a function $F : \mathbb{R} \to \mathbb{C}$, such that*

$$\hat{G}(\omega_1, \omega_2) = F\left(\sqrt{\omega_1^2 + \omega_2^2}\right) \quad \forall -\pi < \omega_1, \omega_2 \le \pi.$$

Isotropic filters are popular, because they treat image features similarly regardless of their directions. Let $\mathrm{grad}(\cdot)$ be the gradient operator. The following theorem, proved in Section 4.1, provides the decay rate of the risk of linear convolution.

Theorem 2. *Consider the linear convolution filter (5) and suppose that g is real, symmetric, and isotropic. Furthermore, assume that $\|\mathrm{grad}(\hat{G}(w_1, w_2))\|_2 \le C$ for a fixed constant C. Then,*

$$\inf_g \sup_{f \in H^\alpha(C)} R_n(f, \hat{f}_g^{LF}) \asymp n^{-\frac{2}{3}}.$$

Castro and Donoho [5] have proved a similar result for the special case of the box filter. While the Horizon model used in [5] is slightly different from our model, their proof works in our setting as well.

1.4.2. Yaroslavsky / SUSAN filter

While linear filters are popular in image processing due to their simplicity, they unfortunately blur images with sharp edges. One popular alternative is

---

[4]For simplicity of analysis, we use a periodic extension of y at the image boundaries.

to adapt the weight of each pixel in the average (5) according to the distance between its noisy value and the value of the pixel we aim to estimate. Let $\mathcal{C}_{i,j}^{\Delta_n} \triangleq \{(m,\ell) \mid i - \Delta_n \leq m \leq i + \Delta_n, j - \Delta_n \leq \ell \leq j + \Delta_n\}$ denote the Δ_n-neighborhood of the pixel (i,j). One popular approach for setting the weights is

$$w_{i,j}^Y(m,\ell) = \mathrm{e}^{-\frac{(y_{m,\ell}-y_{i,j})^2}{2\tau^2}},$$

from which we calculate the estimate

$$\hat{f}_{\Delta_n,\tau}^Y(i,j) = \frac{\sum_{m=i-\Delta_n}^{i+\Delta_n} \sum_{\ell=j-\Delta_n}^{j+\Delta_n} w_{i,j}^Y(m,\ell)\, y_{m,\ell}}{\sum_{m=i-\Delta_n}^{i+\Delta_n} \sum_{\ell=j-\Delta_n}^{j+\Delta_n} w_{i,j}^Y(m,\ell)}. \tag{6}$$

Only the pixels in the Δ_n-neighborhood of (i,j) contribute to the estimate of that pixel. This algorithm is called the *Yaroslavsky Filter* (YF) or *SUSAN filter* [6, 7]; slight modifications are known as the *bilateral filter* [8] and *σ-filter* [9].

To calculate the risk of the YF, we consider a slightly different, oracle-based algorithm. Suppose that in setting the weights $w_{i,j}^Y(m,\ell)$ of the YF we have access to the actual (and not the noisy) value of the pixel (i,j). Using this oracle information we can set the weights according to

$$w_{i,j}^{SY}(m,\ell) = \mathrm{e}^{-\frac{(y_{m,\ell}-x_{i,j})^2}{2\tau^2}}.$$

Intuitively, the oracle weights are "less noisy" than the actual filter weights. Plugging these weights into (6) we obtain what we call the *semi-oracle Yaroslavsky filter* (SYF). The following theorem, proved in Section 4.2, shows that, as far as the decay rate is concerned, the SYF's performance is the same as the linear filter and box filter.

Theorem 3. *The risk of SYF algorithm satisfies*

$$\inf_{\tau,\Delta_n} \sup_{f \in H^\alpha(C)} R_n(f, \hat{f}^{SY}) = \Omega(n^{-2/3}).$$

1.4.3. Sparsity based denoising

Another popular class of image denoising methods exploit sparsity in some transform domain via thresholding. Wavelets are often used as the sparsity domain for natural images. Let $\mathcal{W}(y)$ represent the separable two-dimensional wavelet transform of the image, let $\mathcal{IW}$ represent the inverse

wavelet transform, and let $\mathcal{T}$ be the hard thresholding function, i.e., $\mathcal{T}_\theta(x) = x\mathbf{1}_{\{|x|>\theta\}}$. Then wavelet thresholding denoising corresponds to

$$\hat{f}_\theta^W = \mathcal{IW}(\mathcal{T}_\theta(\mathcal{W}(y))).$$

Donoho and Johnstone have proven that $\sup_{f\in H^\alpha(C)} R_n(f, \hat{f}^W) = \Omega(n^{-1})$ [4], [10]. Even though this rate is an improvement over the above algorithms, is still far from the optimal achievable rate of $n^{-\frac{4}{3}}$ for $\alpha = 2$.

This suboptimality spurred the development of other sparsity-inducing transformations, including curvelets [11], wedgelets [4], shearlets [12], and contourlets [13]. Among these transforms, wedgelet denoising provably achieves the optimal rate of $n^{-\frac{4}{3}}$ for $\alpha = 2$ [4]. However, wedgelet denoising performs poorly on textures, which has limited its application in practice to date.

2. Nonlocal means denoising

The YF estimator sets its weights according the noisy pixel values and their spatial vicinity; however neither of these two features are reliable for noisy, edgy images. In contrast, the *nonlocal means* (NLM) algorithm sets its weights according to the proximity of the image patch surrounding each noisy pixel with other patches in the image [14]. Define the δ_n-neighborhood distance $d_{\delta_n}(y_{i,j}, y_{m,\ell})$ between two observations as

$$d_{\delta_n}^2(y_{i,j}, y_{n,p}) = \frac{1}{\rho_n^2} \sum_{m=-\delta_n}^{\delta_n} \sum_{\ell=-\delta_n}^{\delta_n} |y_{i+\ell,j+m} - y_{n+\ell,p+m}|^2 - |y_{i,j} - y_{n,p}|^2,$$

where $\rho_n^2 = (2\delta_n + 1)^2 - 1$. Note that, in contrast to the definition in [14], we have removed the center element $|y_{i,j} - y_{n,p}|^2$ from the summation. Since we assume that $\delta_n \to \infty$ as $n \to \infty$, the effect is negligible on the asymptotic performance. But, as we will see in Section 4, removing the center element simplifies the calculations considerably. NLM uses the neighborhood distances to estimate

$$\hat{f}_{i,j}^N = \frac{\sum_{(m,\ell)\in S} w_{i,j}^N(m,\ell)y_{m,\ell}}{\sum_{(m,\ell)\in S} w_{i,j}^N(m,\ell)}, \tag{7}$$

where $S = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$ and $w_{i,j}(m, \ell)$ is set according to the δ_n-neighborhood distance between $y_{i,j}$ and $y_{m,\ell}$. For the simplicity of

notation, in cases where both the reference pixel (i,j) and the algorithm are obvious from the context, we will omit the superscript and subscript of the weight and use the simplified notation $w_{m,\ell}$ instead of $w_{i,j}^N(m,\ell)$. It is straightforward to verify that $\mathbb{E}(d_{\delta_n}^2(y_{i,j}, y_{m,\ell})) = d_{\delta_n}^2(x_{i,j}, x_{m,\ell}) + 2\sigma^2$, which suggests the following strategy for setting the weights:

$$w_{i,j}^N(m,\ell) = \begin{cases} 1 & \text{if } d_{\delta_n}^2(y_{i,j}, y_{m,\ell}) \leq 2\sigma^2 + t_n, \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where t_n is the *threshold parameter*. Soft/tapered versions of setting the weights have been explored and are often used in practice [14]. However, the above untapered weights capture the essesnse of the algorithm while simplifying the analysis. We postpone the discussion of tapered weights until Section 5.

There are two main differences between the NLM and YF algorithms. First, the pixels that contribute in the NLM averaging are not necessarily in the local neighborhood of the reference pixel (hence the monicker "nonlocal"). Second, the NLM weights depend not on the difference between the pixel values but on distance between the pixel neighborhoods. In other words the pixel neighborhood is even more important than the pixel value.

To derive a lower bound for the risk of NLM, we will analyze two algorithms that set the weights using some degree of oracle information regarding the true value of the signal. The *full oracle* NLM (FNLM) has access to $\mathbb{E}(d_{\delta_n}^2(y_{i,j}, y_{m,\ell}))$ in setting the weights $w_{m,l}$ in (7) and thus sets them using the noise-free values of the pixels

$$w_{i,j}^F(m,\ell) = \begin{cases} 1 & \text{if } d_{\delta_n}^2(x_{i,j}, x_{m,\ell}) \leq t_n, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The *semi-oracle* NLM (SNLM) differs only slightly from the standard NLM in that it uses the semi-oracle neighborhood distance

$$\bar{d}_{\delta_n}^2(y_{i,j}, y_{n,p}) \triangleq \frac{1}{\rho_n^2} \left(\sum_{m=-\delta_n}^{\delta_n} \sum_{\ell=-\delta_n}^{\delta_n} |x_{i+\ell,j+m} - y_{n+\ell,p+m}|^2 - (x_{i,j} - y_{n,p})^2 \right), \tag{10}$$

and then sets the weights in (7) according to

$$w_{i,j}^S(m,\ell) = \begin{cases} 1 & \text{if } \bar{d}_{\delta_n}^2(y_{i,j}, y_{m,\ell}) \leq \sigma^2 + t_n, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Unlike FNLM, SNLM assumes that just one-half of the noise is removed from the distance estimates. Therefore, the distances calculated in the SNLM are more accurate than the standard NLM but less accurate than in the FNLM. In the rest of the paper, we will use $\hat{f}^N$, $\hat{f}^S$, and $\hat{f}^F$ to denote the NLM, SNLM, and FNLM estimators, respectively.

3. Main Results

Our first result, proved in Section 4.3, establishes an *upper bound* on the risk of NLM.

Theorem 4. *Fix $\epsilon > 0$ and consider NLM denoising with $\delta_n = 2\log^{\frac{1}{2}+\epsilon} n$ and $t_n = \frac{2\sigma^2}{\log^{\frac{\epsilon}{2}} n}$. The risk of this algorithm over the class $H^\alpha(C)$ is*

$$\sup_{f \in H^\alpha(C)} R(f, \hat{f}^N) = O\left(\frac{\log^{\frac{1}{2}+\epsilon} n}{n}\right). \tag{12}$$

Before we discuss the implications of this theorem, it is important to note that, while we can improve the decay rate as close as we desire to $O(n^{-1}\log^{\frac{1}{2}} n)$, the constants that are involved in the big-O notation grow as ϵ decreases. Therefore, in practice very small values of ϵ are not desirable.

Comparing the upper bound (12) with the optimal minimax risk (4) indicates that NLM is *suboptimal* for $\alpha > 1$. In other words, NLM cannnot exploit the smoothness of edge contours in images.

The bound in Theorem 4 is for a specific choice of parameters, and it is natural to ask whether NLM can achieve the optimal rate with some other choice of parameters. To answer this question, we consider SNLM, which outperforms standard NLM in general. We make the following mild assumptions:

A1: The window size $\delta_n \to \infty$ as $n \to \infty$. This assumption is critical to ensuring good performance of any NLM estimator.

A2: The threshold is set to $\sigma^2 + t_n$ as explained in (11) with $t_n > 0$. This ensures that if the neighborhood of pixel (m, ℓ) is exactly the same as the neighborhood of pixel (i, j), then $w_{m,\ell} = 1$ with high probability.

A3: The threshold t_n is set such that, if the noise-free neighborhoods are different in more than half of their pixels, i.e., if $d^2(x_{i,j}, x_{m,\ell}) \geq \frac{1}{2}$, then $\mathbb{P}(w_{i,j}^F(m, \ell) = 1) = o(n^{-1})$.

A4: $\delta_n = O(n^\beta)$, for some $\beta \leq 0.3$.

The following theorem provides a *lower bound* on the performance of SNLM.

Theorem 5. *Suppose that δ_n and t_n satisfy A1–A4. The risk of the SNLM over the class $H^\alpha(C)$ is*

$$\inf_{\delta_n, t_n} \sup_{f \in H^\alpha(C)} R(f, \hat{f}^S) = \Omega(n^{-1}).$$

This bound is still suboptimal compared to the $n^{-4/3}$ minimax rate for $\alpha = 2$. In the words of John Cornyn III, the junior United States Senator for Texas, "The problem with a mini-deal is we have a maxi-problem" [15].

Remarkably, this lower bound is achieved on a very simple image on which NLM would be assumed to work very well: $\mathbf{1}_{\{t_2 < 0.5\}}$ (see Figure 2). Here is what goes wrong. Consider the estimation of an "edge" pixel (i, j) that satisfies $j = \lceil nh(\frac{i}{n}) \rceil$. Define the set $J = \{(m, \ell) \mid \ell = \lfloor nh(\frac{m}{n}) \rfloor\}$ as the set of pixels just below the edge. We will prove the probability that a pixel in J contributes to the NLM estimate ($w_{i,j}(m, \ell) = 1$) is larger than p_0, where p_0 does not depend on n. This happens due to the low "signal to noise ratio" in the distance estimates. Hence $\Theta(n)$ pixels of J will contribute to the NLM estimate. Since these pixels have $x_{m,\ell} = 1$, they introduce a large bias in the estimate. In fact, we show below that the bias, as defined in (2), will be larger than $\frac{np_0}{n + np_0 + np_0}$. Here np_0 corresponds to the pixels below the edge that pass the threshold. This shows that the bias is clearly $\Theta(1)$. Since there are n edge pixels, the risk of the estimator over the entire image is $\Omega(n^{-1})$.

4. Proofs of the Main Theorems

4.1. Proof of Theorem 2

The proof has two main steps. The first step is to prove that there exists a linear filter for which the supremum risk is upper bounded by $O(n^{-2/3})$. For this step we use Theorem 3.1 and 3.2 from [5], which establish the same upper bound for the box filter. The second and more challenging step is to prove that no other linear filter can improve on this decay rate. The rest of this section is dedicated to the proof of this fact.
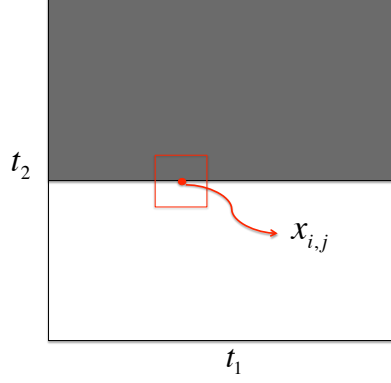
10

Figure 2: The simple image Horizon $\mathbf{1}_{\{t_2<0.5\}}$ used for proving the various lower bounds.

Consider the function $f_h(t_1, t_2)$ for $h(t) = \frac{1}{2}$ and suppose that n is even. This function is displayed in Figure 2. Let

$$X(k_1, k_2) = \frac{1}{n} \sum_{\ell_1} \sum_{\ell_2} x_{\ell_1, \ell_2} e^{-j\frac{2\pi k_1 \ell_1}{n}} e^{-j\frac{2\pi k_2 \ell_2}{n}}$$

represent the Discrete Fourier Transform (DFT) of a discrete two-dimensional discrete signal x. Since $y = x + z$, the DFT of $\hat{f}_g^{LF}$ equals

$$\hat{F}_g^{LF}(k_1, k_2) = Y(k_1, k_2) \cdot G(k_1, k_2) = X(k_1, k_2)G(k_1, k_2) + Z(k_1, k_2)G(k_1, k_2),$$

where Z is again iid $N(0, \sigma^2)$. For $f_h(t_1, t_2)$ with $h(t_1) = \frac{1}{2}$, $X(k_1, k_2)$ satisfies

$$X(k_1, k_2) = \begin{cases} 0 & \text{if } k_1 \neq 0, \\ \frac{1-e^{-j\pi k_2}}{1-e^{-j\frac{2\pi k_2}{n}}} & \text{if } k_1 = 0. \end{cases} \tag{13}$$

It is easy to see that $R_n(f, \hat{f}_g^{LF}) = \frac{1}{n^2} \mathbb{E}(\|X - \hat{F}_g^{LF}\|_F^2)$, where $\|Y\|_F^2 \triangleq \sum_{k_1, k_2} |Y(k_1, k_2)|^2$. If we define $B(\hat{f})$ as the bias of the estimator $\hat{f}$, then we have

$$\begin{aligned} \text{B}^2(\hat{f}_g^{LF}) &= \frac{1}{n^2} \sum_{1 \leq k_2 \leq n, \text{ odd}} |1 - G(0, k_2)|^2 \frac{1}{\sin^2 \frac{\pi k_2}{n}} \\ &\geq \frac{1}{n^2} \sum_{1 \leq k_2 \leq n^{2/3}, \text{ odd}} |1 - G(0, k_2)|^2 \frac{1}{\sin^2 \frac{\pi k_2}{n}}. \end{aligned}$$

11

The variance of the estimator is

$$\mathrm{Var}(\hat{f}_g^{LF}) = \frac{1}{n^2} \sum_{k_1,k_2} |G(k_1,k_2)|^2 \sigma^2.$$

We know that

$$\frac{1}{n^2} \sum_{k_1,k_2} |G(k_1,k_2)|^2 = \int \int |\hat{G}(\omega_1,\omega_2)|^2 + O(n^{-1}), \qquad (14)$$

where $\hat{G}$ is the continuous Fourier transform of g and satisfies $\|\mathrm{grad}(\hat{G})\|_2 \leq C$. Since g is isotropic, there exists $F : \mathbb{R} \to C$ such that

$$\hat{G}(\omega_1,\omega_2) = F\left(\sqrt{\omega_1^2 + \omega_2^2}\right).$$

Changing the variables of integration in (14) to polar coordinate radius $\omega_r = \sqrt{\omega_1^2 + \omega_2^2}$ angle θ, we have

$$\int \int |\hat{G}(\omega_1,\omega_2)|^2 \geq 2\pi \int_{r=0}^{2\pi} r|F(r)|^2 dr = 2\pi \int_{\omega_2=0}^{2\pi} \omega_2 |\hat{G}(0,\omega_2)|^2 d\omega_2. \quad (15)$$

Combining (14) and (15) we have

$$\mathrm{Var}(\hat{f}_h^{LF}) = \frac{1}{n^2} \sum_{k_1,k_2} |G(k_1,k_2)|^2 \sigma^2 = \frac{4\pi^2}{n^2} \sum_{k_2} k_2 |G(0,k_2)|^2 \sigma^2 - O(n^{-1}).$$

Summing the lower bounds for the bias and variance of this estimator, we obtain the following lower bound for the risk of linear filtering:

$$R_n(f,\hat{f}^{LF}) = \mathrm{B}^2(\hat{f}_g^{LF}) + \mathrm{Var}(\hat{f}_g^{LF})$$

$$\geq \frac{1}{n^2} \sum_{1 \leq k_2 \leq n^{2/3},\ \mathrm{odd}} |1 - G(0,k_2)|^2 \frac{1}{\sin^2 \frac{\pi k_2}{n}} + \frac{4\pi^2}{n^2} \sum_{k_2} k_2 |G(0,k_2)|^2 \sigma^2 - O(n^{-1})$$

$$= \frac{1}{n^2} \sum_{1 \leq k_2 \leq n^{2/3},\ \mathrm{odd}} |1 - G(0,k_2)|^2 \frac{n^2}{\pi^2 k_2^2} + \frac{4\pi^2}{n^2} \sum_{k_2} k_2 |G(0,k_2)|^2 \sigma^2 - O(n^{-1}).$$

Minimizing the dominant term of the lower bound over the filter weights provides $G^*(0,k_2) = \frac{1}{1 + \frac{4\pi^4 \sigma^2 k_2^3}{n^2}}$ for odd values of k_2 and zero for even values

12

of k_2. To find a lower bound we calculate the bias term with these optimal weights:

$$
\begin{aligned}
\mathrm{B}^2(\hat{f}_{g^*}^{LF}) &= \frac{1}{n^2} \sum_{1 \leq k_2 \leq n^{2/3}, \text{ odd}} |1 - G(0, k_2)|^2 \frac{n^2}{\pi^2 k_2^2} \\
&= \frac{1}{n^2} \sum_{1 \leq k_2 \leq n^{2/3}, \text{ odd}} \left(\frac{4\pi^4 \sigma^2 k_2^3 / n^2}{1 + 4\pi^4 \sigma^2 k_2^3 / n^2} \right)^2 \frac{n^2}{\pi^2 k_2^2} \\
&\geq \frac{1}{n^2} \sum_{1 \leq k_2 \leq n^{2/3}, \text{ odd}} \left(\frac{4\pi^4 \sigma^2 k_2^3 / n^2}{1 + 4\pi^4 \sigma^2} \right)^2 \frac{n^2}{\pi^2 k_2^2} \\
&= \frac{1}{n^4} \left(\frac{4\pi^4 \sigma^2}{1 + 4\pi^4 \sigma^2} \right)^2 \sum_{1 \leq k_2 \leq n^{2/3}, \text{ odd}} k_2^4 \\
&= \left(\frac{4\pi^4 \sigma^2}{1 + 4\pi^4 \sigma^2} \right)^2 \left(\frac{n^{-2/3}}{40} + o(n^{-2/3}) \right).
\end{aligned}
$$

This completes the proof.

4.2. Proof of Theorem 3

In this section, denote the pixel to be estimated as $x_{i,j}$. For clarity we use the notation $w_{m,\ell}$ instead of $w_{i,j}^{SY}(m, \ell)$. We first characterize some of the properties of the SYF weights.

Lemma 1. *Suppose that $x_{i,j} = 0$. If $x_{m,\ell} = x_{i,j}$, then $\mathbb{E}(w_{m,\ell} y_{m,\ell}) = 0$. Furthermore, if $|x_{i,j} - x_{m,\ell}| = 1$, then $\mathbb{E}(w_{m,\ell} y_{m,\ell}) > \frac{\tau}{\sqrt{\sigma^2 + \tau^2}} e^{\frac{-1}{2(\sigma^2 + \tau^2)}}$.*

Proof. The first claim is clear from symmetry. To prove the second claim, we observe that $\mathbb{E}(w_{m,\ell} y_{m,\ell}) = \mathbb{E}(w_{m,\ell} x_{m,\ell}) + \mathbb{E}(w_{m,\ell} z_{m,\ell})$. Since $x_{m,\ell} = 1$, we calculate $\mathbb{E}(w_{m,\ell})$ and $\mathbb{E}(w_{m,\ell} z_{m,\ell})$. It is clear that $\mathbb{E}(w_{m,\ell} z_{m,\ell}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z_{m,\ell} e^{-\frac{(z_{m,\ell} - 1)^2}{2\tau^2} - \frac{z_{m,\ell}^2}{2\sigma^2}} \geq 0$. Therefore we calculate

$$
\begin{aligned}
\mathbb{E}(w_{m,\ell}) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(z_{m,\ell} - 1)^2}{2\tau^2} - \frac{z_{m,\ell}^2}{2\sigma^2}} \\
&= \frac{e^{-\frac{1}{2\tau^2} + \frac{\sigma^2}{2(\sigma^2 + \tau^2)\tau^2}}}{\sigma 2\pi} \int_{-\infty}^{\infty} e^{\frac{-\sigma^2 + \tau^2}{2\sigma^2\tau^2}(z_{m,\ell}^2 - \frac{2\sigma^2}{\sigma^2 + \tau^2} z_{m,\ell} + \frac{\sigma^4}{(\sigma^2 + \tau^2)^2})} \\
&= \frac{e^{-\frac{1}{2(\sigma^2 + \tau^2)}}}{\sigma} \sqrt{\frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2}} = \frac{\tau e^{-\frac{1}{2(\sigma^2 + \tau^2)}}}{\sqrt{\sigma^2 + \tau^2}}.
\end{aligned}
$$

13

This completes the proof. □

Define the Δ-neighborhood of a pixel (m, ℓ) as $\mathcal{C}^{\Delta}_{m,\ell} = \{(i,j) : |i - m| \leq \Delta, |j - \ell| \leq \Delta\} \cap S$.

Lemma 2. *Let $\Omega_n = (2\Delta_n + 1)^2$. We then have*

$$\mathbb{P}\left(\frac{1}{\Omega_n}\left(\sum_{(m,\ell)\in\mathcal{C}^{\Delta_n}_{i,j}} w^{SY}_{m,\ell} - \sum_{(m,\ell)\in\mathcal{C}^{\Delta_n}_{i,j}} \mathbb{E}w^{SY}_{m,\ell}\right) \geq t\right) \leq 2e^{-2\Omega_n t^2}.$$

The proof is a simple application of the Hoeffding inequality.

Proof of Theorem 3. The first claim is that the optimal neighborhood size satisfies $\Delta_n = \Omega(\log n)$. We prove this by contradiction. Suppose that $\Delta_n = O(\log(n))$ and consider the performance of the SYF on the image $x_{i,j} = 0$ for every (i,j). It is clear that the bias is zero. However, the variance is lower bounded by $\Omega\left(\frac{1}{\log^2 n}\right)$. This is far from the optimal performance of the linear filters analyzed in Theorem 2. Therefore $\Delta_n = \Omega(\log(n))$.

Now consider the example image shown in Figure 2 with $f_h(t_1, t_2) = \mathbf{1}_{\{t_2 < 0.5\}}$. For notational simplicity we assume n is even so that the value of each pixel is either 0 or 1. Define the two regions $P_1 = \{(i,j) : \frac{n}{2} \leq j \leq \frac{n}{2} + \frac{\Delta_n}{2}\}$ and $P_2 = \{(i,j) : j > \frac{n}{2} + \Delta_n\}$. At least 1/4 of the pixels in the neighborhood of the pixels in P_1 have the noise-free value of 1. All pixels in the neighborhood of the pixels in P_2 have the noise-free pixel values equal to 1. Over each region we will find a lower bound for the risk of SYF and then sum them to obtain a lower bound for the risk over the entire image.

Case I $- (i,j) \in P_1$: From the Jensen inequality we have

$$\mathbb{E}\left(x_{i,j} - \frac{\sum_{(m,\ell)\in\mathcal{C}^{\Delta_n}_{i,j}} w_{m,\ell}y_{m,\ell}}{\sum_{(m,\ell)\in\mathcal{C}^{\Delta_n}_{i,j}} w_{i,j}}\right)^2 \geq \left(\mathbb{E}\frac{\sum_{(m,\ell)\in\mathcal{C}^{\Delta_n}_{i,j}} w_{m,\ell}y_{m,\ell}}{\sum_{(m,\ell)\in\mathcal{C}^{\Delta_n}_{i,j}} w_{m,\ell}}\right)^2.$$

Define the following two constants:

$$m_0 = \mathbb{E}(w^{SY}_{i,j}(m, \ell) \mid x_{i,j} = 0, x_{m,\ell} = 0),$$
$$m_1 = \mathbb{E}(w^{SY}_{i,j}(m, \ell) \mid x_{i,j} = 0, x_{m,\ell} = 1).$$

14

It is clear that $m_0 > m_1$. Let the event A be

$$A = \left\{ \sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} - \sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} \mathbb{E} w_{m,\ell} \leq \Delta_n^{2-\epsilon} \right\} \tag{16}$$

for some $\epsilon > 0$. We have

$$\mathbb{E} \left(\frac{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} y_{m,\ell}}{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell}} \right) \geq \mathbb{E} \left(\frac{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} y_{m,\ell}}{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell}} \,\middle|\, A \right) \mathbb{P}(A)$$

$$\stackrel{(a)}{\geq} \mathbb{E} \left(\frac{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} y_{m,\ell}}{4\Delta_n^2 m_0 + \Delta_n^{2-\epsilon}} \,\middle|\, A \right) \mathbb{P}(A)$$

$$\geq \mathbb{E} \left(\frac{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} y_{m,\ell}}{4\Delta_n^2 m_0 + \Delta_n^{2-\epsilon}} \right) - \mathbb{P}(A^c)$$

$$\stackrel{(b)}{\geq} \left(\frac{\Delta_n^2 c_0}{4\Delta_n^2 m_0 + \Delta_n^{2-\epsilon}} \right) - P(A^c).$$

Inequality (a) uses Lemma 2 and the fact that $m_0 \geq m_1$. Inequality (b) uses Lemma 1, and therefore $c_0 = \frac{\tau}{\sqrt{\sigma^2 + \tau^2} e^{-\frac{1}{2(\sigma^2 + \tau^2)}}}$. Since $\mathcal{C}_{i,j}^{\Delta_n}$ has $(2\Delta_n + 1)^2$ pixels, at least Δ_n^2 of them have the noise-free pixel value 1.

Since $\Delta_n = \Omega(\log n)$, Lemma 2 proves that $P(A^c) = o(1)$ and, therefore, the bias is lower bounded by $\Theta(1)$ for all of the pixels in P_1.

Case II $- (i,j) \in P_2$: As mentioned before, all pixels in the neighborhood of the pixels in P2 have the noise-free pixel values equal to 1. Hence, we have

$$\mathbb{E} \left(\frac{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} y_{i,j}}{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell}} \right)^2 = \mathbb{E} \left(\frac{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell) \in \mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell}} \right)^2.$$

Defining the event A as in (16), we have

$$\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in\mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in\mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell}}\right)^2 \middle| A\right)\mathbb{P}(A)$$

$$\geq \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in\mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} z_{m,\ell}}{4\Delta_n^2 m_0 + \Delta_n^{2-\epsilon}}\right)^2 \middle| A\right)\mathbb{P}(A)$$

$$\geq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in\mathcal{C}_{i,j}^{\Delta_n}} w_{m,\ell} z_{m,\ell}}{4\Delta_n^2 m_0 + \Delta_n^{2-\epsilon}}\right)^2 - \mathbb{P}(A^c) = \frac{4\Delta_n^2\mathbb{E}(w_{m,\ell} z_{m,\ell})^2}{(4m_0\Delta_n^2 + \Delta_n^{2-\epsilon})^2} - \mathbb{P}(A^c).$$

If the neighborhood size is larger than $c\log(n)$ for some constant c, then Lemma 2 will imply that $\mathbb{P}(A^c) < o\left(\frac{1}{n^2}\right)$. Therefore, the dominant term in the above expression of the form of $\frac{\gamma}{\Delta_n^2}$. Combining the lower bounds for P_1 and P_2, we obtain a lower bound of the form of $\frac{\beta\Delta_n}{n} + \frac{\gamma}{\Delta_n^2}$. Optimizing over Δ_n proves that

$$\inf_{\Delta_n,\tau} R_n(f, \hat{f}^{SY}) > \Omega(n^{-2/3}).$$

This completes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

It is clear from the proof above that the neighborhood size is the main parameter that controls the decay rate of the risk of the YF. The Gaussian term in the YF weights enables an improvement in the constants but does not play any role in the decay rate. In the extreme case of $\Delta_n = n$, when all of the image pixels can potentially contribute to the estimation of a pixel, the decay rate of YF degrades to $\Theta(1)$. This algorithm is called the *range filter*, and [8] observed in practice that it performs much worse than even linear filters, as the above analysis confirms. Interestingly, NLM addresses this issue and therefore its search space could be the entire image. This is the main reason for its improved performance.

The lower bound proved in Theorem 3 is the same as the upper bound we derived for the performance of linear filtering. Therefore, we have the following theorem.

Theorem 6. *The risk of the SYF satisfies*

$$\inf_{\Delta_n,\tau} \sup_{f\in H^\alpha(C)} R_n(f, \hat{f}^{SY}) \asymp n^{-2/3}.$$

16

4.3. Proof of Theorem 4

The proof has two main steps. First, we show that the risk of the pixels far from the edge is $O(\log^{1+2\epsilon}(n)/n^2)$. Second, we show that the risk of the pixels whose δ_n neighborhood senses the edge is constant; however there are at most $O(n\delta_n)$ of these pixels. The following two lemmas will play key roles in our analysis.

Lemma 3. *Let $Z \sim N(0, \sigma^2)$. For $\lambda < \frac{1}{2\sigma^2}$, we have*

$$\mathbb{E}(e^{\lambda Z^2}) = \frac{1}{\sqrt{1 - 2\lambda\sigma^2}}.$$

Proof. The proof is a simple integral calculation:

$$\mathbb{E}(e^{\lambda Z^2}) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(\lambda - \frac{1}{2\sigma^2})Z^2} dZ = \frac{1}{\sigma\sqrt{\frac{1}{\sigma^2} - 2\lambda}}.$$

$\square$

Lemma 4. *Let $Z_1, Z_2, \ldots, Z_n$ be iid $N(0,1)$ random variables. The χ_n^2 random variable defined as $\sum_{i=1}^{n} Z_i^2$ concentrates around its mean with high probability, i.e.,*

$$\mathbb{P}\left(\frac{1}{n}\sum_i Z_i^2 - 1 > t\right) \leq e^{-\frac{n}{2}(t - \ln(1+t))},$$

$$\mathbb{P}\left(\frac{1}{n}\sum_i Z_i^2 - 1 < -t\right) \leq e^{-\frac{n}{2}(t + \ln(1-t))}.$$

Proof. Here we prove just the first claim; the proof of the second claim follows along very similar lines. From Markov's Inequality, we have

$$\mathbb{P}\left(\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^2\right) - 1 > t\right) \leq e^{-\lambda t - \lambda}\mathbb{E}\left(e^{\frac{\lambda}{n}\sum_{i=1}^{n} Z_i^2}\right)$$

$$= e^{-\lambda t - \lambda}\left(\mathbb{E}\left(e^{\frac{\lambda Z_1^2}{n}}\right)\right)^n = \frac{e^{-\lambda t - \lambda}}{\left(1 - \frac{2\lambda}{n}\right)^{\frac{n}{2}}}. \tag{17}$$
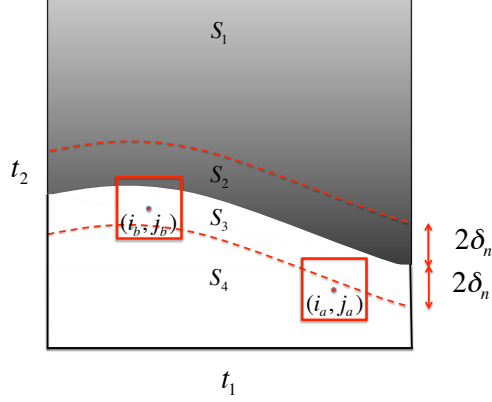
17

Figure 3: An example of a Horizon image. The δ_n-neighborhood of pixel $(i_a, j_a) \in S_4$ does not intersect the edge contour, while the δ_n-neighborhood of pixel $(i_b, j_b) \in S_3$ intersects with the edge contour.

The last inequality follows from Lemma 3. The upper bound proved above holds for any $\lambda < \frac{n}{2}$. To obtain the lowest upper bound we minimize $\frac{\mathrm{e}^{-\lambda t - \lambda}}{\left(1 - \frac{2\lambda}{n}\right)^{\frac{n}{2}}}$ over λ. The optimal value of λ is $\lambda^\star = \arg\min_\lambda \frac{\mathrm{e}^{-\lambda t - \lambda}}{\left(1 - \frac{2\lambda}{n}\right)^{\frac{n}{2}}} = \frac{nt}{2(t+1)}$. Plugging λ^* into (17) proves the result. $\qquad\square$

Proof of Theorem 4. We will consider the following partition of the image pixels. Let $S = \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$. For a given Horizon function $f_h(t_1, t_2)$, define $S_1 = \{(i,j) \mid \frac{j}{n} > h(\frac{i}{n}) + \frac{2\delta_n}{n}\}$, $S_2 = \{(i,j) \mid h(\frac{i}{n}) < \frac{j}{n} \leq h(\frac{i}{n}) + \frac{2\delta_n}{n}\}$, $S_3 = \{(i,j) \mid h(\frac{i}{n}) - \frac{2\delta_n}{n} \leq \frac{j}{n} \leq h(\frac{i}{n})\}$, and $S_4 = \{(i,j) \mid \frac{j}{n} < h(\frac{i}{n}) - \frac{2\delta_n}{n}\}$. These regions are displayed in Figure 3. The δ_n-neighborhood of the pixels in S_1 and S_4 do not intersect the edge, while the δ_n-neighborhood of the other pixels may have pixels from both sides of the edge. See Figure 3. For the notational simplicity we write $\sum_{(i,j) \in S_\ell}$ for the double summation over i, j where j satisfies the constraints specified for S_ℓ.

Consider a pixel $(i, j) \in S_1$. The risk of NLM at this pixel is

$$\mathbb{E}\left(x_{i,j} - \frac{\sum w_{m,\ell} y_{m,\ell}}{\sum w_{m,\ell}}\right)^2,$$

18

where $x_{i,j} = 0$, since $(i, j) \in S_1$. Define the set of oracle weights

$$w^\star_{m,\ell} = \begin{cases} 1 & \text{if } \frac{\ell}{n} > h(\frac{m}{n}), \\ 0 & \text{otherwise.} \end{cases} \qquad (18)$$

Define $U \triangleq \left(\frac{\sum w_{m,\ell} y_{m,\ell}}{\sum w_{m,\ell}} \right)^2$, and let the event $A = \{ w_{m,\ell} = w^\star_{m,\ell}, \ \forall (m, \ell) \in S_1 \cup S_4 \}$. We then have

$$\mathbb{E}(U) = \mathbb{E}(U \mid A)\mathbb{P}(A) + \mathbb{E}(U \mid A^c)\mathbb{P}(A^c) \leq \mathbb{E}(U \mid A)\mathbb{P}(A) + \mathbb{P}(A^c), \quad (19)$$

where the last inequality is due to the fact that the risk of the estimator is bounded by 1. We now calculate each term of (19) separately.

Define $S_{14} = S_1 \cup S_4$ and $S_{23} = S_2 \cup S_3$. Then we have

$$\mathbb{E}(U \mid A)\mathbb{P}(A)$$
$$= \ \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} y_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} y_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2 \middle| A \right) \mathbb{P}(A)$$
$$\leq \ \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} y_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} y_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2$$
$$\leq \ \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2$$
$$+ \ \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2$$
$$+ \ 2\sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2}$$
$$\times \ \sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_1 \cup S_4} w^\star_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^\star_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2}. \qquad (20)$$

The last inequality is due to Cauchy-Schwartz. In the next two lemmas we bound the last three terms of (20).

19

Lemma 5. *Let $w_{m,\ell}$ be the weights of NLM with $\delta_n = \log^{\frac{1}{2}+\epsilon} n$ and $t_n = \frac{2}{\sqrt{\log^{\frac{\epsilon}{2}} n}}$ for $\epsilon > 0$. Also, let $w^{\star}_{m,\ell}$ be the oracle weights introduced in (18). Then*

$$\mathbb{E} \left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2 = O \left(\frac{\delta_n^2}{n^2} \right).$$

Proof. Define S_f as the set of indices of the pixels whose noise-free value is neither zero nor one. Since the images are chosen from the Horizon class, the cardinality of this set is at most $2n$. Plugging in the values of $x_{m,\ell}$, we have

$$\mathbb{E} \left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2$$

$$\overset{(a)}{=} \mathbb{E} \left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_3 \backslash S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} + \sum_{(m,\ell)\in S_3 \backslash S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_2 \backslash S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m\ell}} \right)^2$$

$$\overset{(b)}{\leq} \mathbb{E} \left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_3} 1 + \sum_{(m,\ell)\in S_f} w_{m\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} + \sum_{(m,\ell)\in S_3} 1 + \sum_{(m,\ell)\in S_f} w_{m,\ell}} \right)^2$$

$$\leq \mathbb{E} \left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_3} 1 + 2n}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} + \sum_{(m,\ell)\in S_3} 1} \right)^2 = O \left(\frac{\delta_n^2}{n^2} \right),$$

where Inequality (b) is due to the fact that the expression after Equality (a) is an increasing function of $\sum_{(m,\ell)\in S_3 \backslash S_f} w_{m\ell}$ and a decreasing function of $\sum_{(m,\ell)\in S_2 \backslash S_f} w_{m\ell}$. Therefore, we set $w_{m,\ell} = 1$ for $(m,\ell) \in S_3$ and $w_{m,\ell} = 0$ for $(m,\ell) \in S_2$. $\qquad\square$

Lemma 6. *Let $w_{m,\ell}$ be the weights of NLM with $\delta_n = \log^{\frac{1}{2}+\epsilon} n$ and $t_n = \frac{2}{\sqrt{\log^{\frac{\epsilon}{2}} n}}$ for $\epsilon > 0$. Also, let $w^{\star}_{m,\ell}$ be the oracle weights introduced in (18). Then we have*

$$\mathbb{E} \left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}} \right)^2 = O \left(\frac{1}{n^2} \right).$$

Proof. Since $\sum_{(m,\ell)\in S_{23}} w_{m,\ell} \geq 0$ and we are interested in the upper bound

of the risk, we can remove it from the denominator to obtain

$$\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell}}\right)^2\right)$$

$$\leq \quad \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell} + \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2\right)$$

$$= \quad \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2\right) + \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2\right)$$

$$+ \; 2\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)\right). \qquad (21)$$

Since $\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}$ is the average of iid random variables, it is not hard to prove that $\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2 = O(\frac{\sigma^2}{n^2})$. To bound the other two terms in (21) we use the notation defined in the last section: $\mathcal{C}^{\Delta}_{m,\ell} = \{(i,j) : |i - m| < \Delta, |j - \ell| < \Delta\} \cap S$. We also define $\mathbb{E}(\cdot \mid \mathcal{C}^{\Delta}_{m,\ell})$ as the conditional expectation given the variables in $\mathcal{C}^{\Delta}_{m,\ell}$. We then have

$$\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2\right)$$

$$= \quad \mathbb{E}\left(\mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell}}\right)^2 \;\middle|\; \mathcal{C}^{\delta_n}_{i,j}\right)\right)$$

$$= \quad \mathbb{E}\left(\frac{\mathbb{E}(\sum_{(m',\ell')\in S_{23}} \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'} \mid \mathcal{C}^{\delta_n}_{i,j})}{(\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell})^2}\right)$$

$$= \quad \mathbb{E}\left(\frac{\mathbb{E}(\sum_{(m',\ell')\in \mathcal{C}^{2\delta_n}_{m,\ell}} \sum_{(m,\ell)\in S_{23}} w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'} \mid \mathcal{C}^{\delta_n}_{i,j})}{(\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell})^2}\right)$$

$$= \quad \left(\frac{\sum_{(m',\ell')\in \mathcal{C}^{2\delta_n}_{m,\ell}} \sum_{(m,\ell)\in S_{23}} \mathbb{E}(w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'})}{(\sum_{(m,\ell)\in S_{14}} w^{\star}_{m,\ell})^2}\right) \leq O\left(\frac{\delta_n^3}{n^3}\right).$$

For the last inequality we have used the Cauchy-Schwartz Inequality to prove

21

that $\mathbb{E}(w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'}) \leq 3\sigma^2$. Although we could derive a loose bound for $\mathbb{E}\left(\left(\frac{\sum_{(m,\ell) \in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w_{m,\ell}^{\star}}\right)^2\right)$ and still draw the same conclusion, we used the above technique since we have to use it in the proof of Theorem 7. The last term we have to bound in (21) is

$$\mathbb{E}\left(\left(\frac{\sum_{(m,\ell) \in S_{14}} w_{m,\ell}^{\star} z_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w_{m,\ell}^{\star}}\right)\left(\frac{\sum_{(m,\ell) \in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w_{m,\ell}^{\star}}\right)\right)$$

$$\leq \sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell) \in S_{14}} w_{m,\ell}^{\star} z_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w_{m,\ell}^{\star}}\right)^2}\sqrt{\mathbb{E}\left(\frac{\sum_{(m,\ell) \in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w_{m,\ell}^{\star}}\right)^2}$$

$$\leq O\left(\frac{1}{n^2}\right).$$

This proves the lemma. $\qquad \square$

Using Lemma 5 and Lemma 6 in (20) proves that

$$\mathbb{E}(U \mid A)\mathbb{P}(A) = O\left(\frac{\delta_n^2}{n^2}\right). \tag{22}$$

Finally, using Lemma 4 and the union bound it is easy to show that

$$\mathbb{P}(A^c) = O\left(\frac{1}{n^2}\right). \tag{23}$$

It is important to note that the constants of this probability are hidden in the O notation. These constants depend on ϵ and increase as ϵ decreases. Therefore, we cannot set $\epsilon = 0$.

Plugging in (23) and (22) in (19) results in

$$\mathbb{E}\left(x_{i,j} - \frac{\sum w_{m,\ell} y_{m,\ell}}{\sum w_{m,\ell}}\right)^2 = O\left(\frac{\log^{1+2\epsilon}(n)}{n^2}\right) \qquad \forall (i,j) \in S_1.$$

Now consider $(i,j) \in S_2 \cup S_3$. In this region we can bound the error by the worst possible risk, which is 1. We will discuss the sharpness of this bound in the next section where we develop a lower bound for the risk.

Using the bounds provided above for the risks of the pixels in S_1, S_2, S_3 and S_4, we can now calculate the final upper bound for the risk of the NLM

as

$$\sup_{f \in H^\alpha(C)} R(f, \hat{f}^{NL}) = \frac{1}{n^2} \sum_i \sum_j \mathbb{E}(x_{i,j} - \hat{f}^N_{i,j})^2$$

$$\leq \frac{\log^{1+2\epsilon}(n)(|S_1| + |S_4|)}{n^2} + \frac{|S_2| + |S_3|}{n^2}$$

$$\leq O\left(\frac{\log^{\frac{1}{2}+\epsilon}(n)}{n}\right).$$

In order to derive the last inequality we noted that since $h(t_1) \in H\ddot{o}lder^1(1)$ the cardinality of S_2 and S_3 are $O(n \log(n))$. This completes the proof of Theorem 4. $\square$

4.4. Proof of Theorem 5

Suppose that the parameters of SNLM satisfy assumptions A1–A4. To derive a lower bound we consider the performance of the SNLM algorithm on the simple image in Figure 2. For notational simplicity we assume that n is even, and hence all of the pixel values are either 0 or 1. The proof follows four main steps:

1. We consider the pixels that are just above the edge, i.e., $(i, \lceil \frac{n}{2} \rceil)$, and prove that the risk of the NLM on these pixels is lower bounded by a constant that does not depend on n.
2. Using asymptotic arguments we prove that the probability a pixel just below the edge passes the threshold $t_n > 0$ is larger than p_0, where p_0 is a non-zero probability independent of n. Based on this, we use a concentration argument to prove that $\Theta(n)$ of the pixels just below the edge will pass the threshold with high probability. See the formal statement in Theorem 1.
3. Using symmetry arguments we prove that the probability a pixel that is $\ell < \delta_n/2$ rows[5] above the edge or below the edge passes the threshold is equal. This is formally stated in Lemma 8.
4. Combining the outcomes of Steps 2 and 3 we show that the risk is minimized if all the pixels just above the edge pass the threshold and the probability that the other pixels pass the threshold is as low as

---

[5]The ℓth *row* of an image is the set of all pixels of the form (i, ℓ).

possible. If more zero pixels above the edge pass the threshold, then more pixels with noise-free value 1 will also pass the threshold, and this makes the bias large. Therefore we assume that $p_{n,\ell}$, the probability that a pixel at distance ℓ of the edge passes the threshold, is equal to zero for $\ell > 1$. However, we have already proven that for $\ell = 1$ the probability is larger than p_0. Theorem 5 uses this fact to show that the risk of this estimator is larger than a constant independent of n.

Proposition 1. *Let* $j^* = \lceil \frac{n}{2} \rceil$. *For any pixel with coordinates of the form* (i^*, j^*), *there exists a non-zero constant probability* p_0 *such that for any* δ_n *and* t_n

$$\mathbb{P}\left(\sum_m w_{m,j^*-1} - np_0 < -t \right) \leq 4\delta_n e^{-\frac{t^2}{4n\delta_n}}.$$

Proof. For notational simplicity we use $i = i^*$ and $j = j^*$ in the proof. We have

$$\mathbb{P}(\bar{d}^2_{\delta_n}(y_{i,j}, y_{m,j-1}) \leq \sigma^2 + t_n)$$

$$= \mathbb{P}\left(\frac{1}{\rho_n^2}(\sum_{\ell,p} |x_{i+p,j+\ell} - y_{m+p,j-1+\ell}|^2 - (x_{i,j} - y_{p,j-1})^2) \leq \sigma^2 + t_n \right)$$

$$= \mathbb{P}\left(\frac{1}{\rho_n^2} \sum_{\ell,p}(s^2_{\ell,p} - \sigma^2) - \frac{2}{\rho_n^2} \sum_{\ell} s_{\ell,0} \leq -\frac{1}{\rho_n} + t_n \right)$$

$$\geq \mathbb{P}\left(\frac{1}{\rho_n^2} \sum_{\ell,p}(s^2_{\ell,p} - \sigma^2) - \frac{2}{\rho_n^2} \sum_{\ell} s_{\ell,0} \leq -\frac{1}{\rho_n} \right),$$

where $s_{\ell,m} = z_{m+\ell,j-1+p}$. According to the Berry-Esseen Central Limit Theorem for independent non-identically distributed random variables [16], we know that

$$\mathbb{P}\left(\frac{1}{\rho_n^2} \sum_{\ell} \sum_{p}(s^2_{\ell,p} - \sigma^2) - \frac{2}{\rho_n^2} \sum_{\ell} s_{\ell,0} \leq -\frac{1}{\rho_n} \right) \geq \mathbb{P}(G \leq -1) - \frac{C}{\rho_n},$$

where G is a Gaussian random variable with mean zero and bounded standard deviation. In fact, it is not difficult to confirm that

$$\mathbb{E}(G^2) = 2\sigma^4 + \frac{8\sigma^2\delta_n - 2\sigma^4}{(2\delta_n + 1)^2}.$$

Since $\mathbb{P}(G \leq -1) \geq 2p_0$ ($2p_0$ is $P(G' \leq -1)$ where $G' \sim N(0, 2\sigma^4)$) is non-zero, for large values of n we can ensure that $C/n < p_0$ and therefore that $\mathbb{P}(\bar{d}^2_{\delta_n}(y_{i,j}, y_{m,j-1}) \leq \sigma^2 + t_n) > p_0$. We now prove that even though the weights are correlated, $\Theta(n)$ of the weights will be equal to 1 with very high probability. Define u_i as $w_{i,j-1}$ and define the process $U = (u_1, \ldots, u_n)$. Break this sequence into $2\delta_n$ subsequences $U_i = (u_i, u_{i+2\delta_n}, u_{i+4\delta_n}, \ldots, u_{n-2\delta_n+i})$. Each U_i has independent and identically distributed elements. Therefore, according to the Hoeffding Inequality, we have $\mathbb{P}(|\sum_{u_j \in U_i} u_j - \frac{n}{2\delta_n}\mathbb{E}(u_i)| > t) \leq 2\mathrm{e}^{\frac{-t^2\delta_n}{n}}$. On the other hand we know that $E(u_i) > p_0$. Therefore,

$$\mathbb{P}\left(\sum_{u_j \in U_i} u_j < \frac{n}{2\delta_n}p_0 - t\right) \leq 2\mathrm{e}^{\frac{-t^2\delta_n}{n}}.$$

Finally we use the union bound to obtain

$$\mathbb{P}\left(\sum u_i - np_0 \leq -t\right) \leq \mathbb{P}\left(\sum_i \sum_{u_j \in U_1} u_j - \frac{n}{2\delta_n}p_0 \leq -t\right)$$

$$\leq \mathbb{P}\left(\cup_i\{\omega : \sum_{u_j \in U_i} u_j - \frac{n}{2\delta_n}p_0 \leq -\frac{t}{2\delta_n}\}\right) \leq 4\delta_n\mathrm{e}^{-\frac{t^2}{4n\delta_n}}.$$

$\square$

Define the set $J = \{(i, j) \mid j = \lfloor jh(\frac{i}{n})\rfloor\}$. It is clear that $|J| = n$. The following Corrollary to Proposition 1 shows that NLM sets the weights of most of the pixels in J to 1.

Corollary 1. *Consider the image displayed in Figure 2, and let $\delta_n = O(n^\alpha)$ for $\alpha < 1$. For any δ_n and $t_n > 0$, $\Theta(n)$ of the pixels in J will pass the threshold t_n with very high probability.*

Proof. Set $t = n^{\frac{3+\alpha}{4}}$ in Proposition 1. $\square$

Remarkably the above corollary holds in a very general setting even if the assumptions A1–A4 do not hold. In other words, NLM in its most general form is not able to distinguish between the pixels right above the edge from the pixels right below the edge. This is due to the fact that the "signal to noise ratio" in the δ_n-neighborhood distance estimates is too low at the edge pixels. This is the result of the isotropic neighborhoods used to form the weight estimates.

Lemma 7. *If $|m - i^*| > \delta_n/2$ and $|m' - i^*| > \delta_n/2$, then*

$$\mathbb{P}(\bar{d}^2_{\delta_n}(y_{i^*,j^*}, y_{m,j^*-\ell}) \leq \sigma^2 + t_n) = \mathbb{P}(\bar{d}^2_{\delta_n}(y_{i^*,j^*}, y_{m',j^*-\ell}) \leq \sigma^2 + t_n)$$

for any ℓ, m, m'.

The proof of this lemma is obvious and is skipped here.

Lemma 8. *For $\ell < \delta_n/2$,*

$$\mathbb{P}(\bar{d}^2_{\delta_n}(y_{i^*,j^*}, y_{m,j^*-\ell}) \leq \sigma^2 + t_n) = \mathbb{P}(\bar{d}^2_{\delta_n}(y_{i^*,j^*}, y_{m,j^*+\ell}) \leq \sigma^2 + t_n).$$

The proof of this lemma is also obvious from symmetry and is skipped here. We can now prove Theorem 5, which provides a lower bound for the risk of SNLM.

Proof of Theorem 5. We derive a lower bound for the risk of SNLM on the image displayed in Figure 2. To do so, we consider the pixels just above the edge and prove that the SNLM algorithm has risk $\Theta(1)$ at these pixels. Since there are $\Theta(n)$ of these pixels, the risk over the entire image is larger than $\Theta(n^{-1})$.

Consider a pixel (i^*, j^*) with $j^* = \lceil \frac{n}{2} \rceil$. The risk of the SNLM is

$$\mathbb{E}\left(f_{i^*,j^*} - \frac{\sum\sum w_{m,\ell} y_{m,\ell}}{\sum\sum w_{m,l}} \right)^2 \geq \left(\mathbb{E}\left(\frac{\sum\sum w_{m,\ell} y_{m,\ell}}{\sum\sum w_{m,l}} \right) \right)^2. \tag{24}$$

Note that $w_{m,\ell}$ is independent of the $y_{m,\ell}$ according to the construction of the SNLM weights in (10). Let $p_{n,\ell}$ be the probability $\mathbb{P}(w_{\ell,i} = 1)$ for $\ell \in \{j^* - \delta_n, j^* - \delta_n + 1, \ldots, j^* + \delta_n\}$. We can partition the row $\{(i, \ell) \mid 1 \leq i \leq n\}$ into $2\delta_n + 1$ subsequences and apply Hoeffding inequality on each subsequence. We combine the results of different subsequences with the union bound to prove that

$$\mathbb{P}\left(|\sum_m w_{m,\ell} - np_{n,\ell}| > t \right) \leq 4\delta_n e^{\frac{-t^2}{4n\delta_n}}. \tag{25}$$

Define the event A as

$$A = \left\{ |\sum_m w_{m,\ell} - np_{n,\ell}| < n^{0.66} \quad \forall \ell, |\ell - j^*| \leq \delta_n \right\}.$$

26

Using the union bound and (25) we have

$$P(A^c) \le 8\delta_n^2 e^{\frac{-n^{1.32}}{4n\delta_n}}.$$

Any lower bound on the bias of the estimator leads to a lower bound on its risk. Therefore, we find a lower bound for the bias as follows:

$$\mathbb{E}\left(\frac{\sum\sum w_{m,\ell}y_{m,\ell}}{\sum\sum w_{m,\ell}}\right) \ge \mathbb{E}\left(\frac{\sum\sum w_{m,\ell}y_{m,\ell}}{\sum\sum w_{m,\ell}} \,\Big|\, A\right)\mathbb{P}(A)$$

$$\ge \mathbb{E}\left(\frac{\sum\sum w_{m,\ell}y_{m,\ell}}{\sum np_{n,\ell}+n^{.66}\delta_n} \,\Big|\, A\right)\mathbb{P}(A) \ge \mathbb{E}\left(\frac{\sum\sum w_{m,\ell}y_{m,\ell}}{\sum np_{n,\ell}+n^{.66}\delta_n}\right) - P(A^c),$$

where for the last inequality we have used the fact that the risk of SNLM is bounded by 1. Since from the construction of SNLM in (10), $w_{m,\ell}$ is independent of $z_{m,\ell}$, we have

$$\mathbb{E}\left(\frac{\sum\sum w_{m,\ell}y_{m,\ell}}{\sum np_{n,\ell}+n^{0.66}\delta_n}\right) - P(A^c) = \mathbb{E}\left(\frac{\sum\sum w_{m,\ell}x_{m,\ell}}{\sum np_{n,\ell}+n^{0.66}\delta_n}\right) - P(A^c)$$

$$= \frac{\sum_{\ell<j^*}np_{n,\ell}}{\sum np_{n,\ell}+n^{0.66}\delta_n} - P(A^c) \ge \frac{\sum_{\ell<j^*}np_{n,\ell}}{n+2\sum_{\ell<j^*}np_{n,\ell}+n^{0.66}\delta_n} - P(A^c).$$

Proposition 1 proves that both the numerator $\sum_{\ell<j^*}np_{n,\ell}$ and the denomenator $\sum np_{n,\ell}+n^{0.66}\delta_n$ are $\Omega(n)$. Therefore, according to $A3$, we can ignore the summations $\sum_{\ell<j^*-\frac{\delta_n}{2}}np_{n,\ell}$ and $\sum_{\ell>j^*+\frac{\delta_n}{2}}np_{n,\ell}$. By combining this fact with Lemma 8, we obtain

$$\frac{\sum_{\ell<j^*}np_{n,\ell}}{\sum np_{n,\ell}+n^{0.66}\delta_n} - P(A^c) \ge \frac{\sum_{\ell<j^*}np_{n,\ell}}{np_{n,j^*}+2\sum_{\ell<j^*}np_{n,\ell}+n^{0.66}\delta_n} - P(A^c)$$

$$\ge \frac{\sum_{\ell<j^*}np_{n,\ell}}{n+2\sum_{\ell<j^*}np_{n,\ell}+n^{0.66}\delta_n} - P(A^c).$$

In the last inequality we assumed that $p_{n,j^*}=1$. To find a lower bound for $\frac{\sum_{\ell<j^*}np_{n,\ell}}{n+2\sum_{\ell<j^*}np_{n,\ell}+n^{0.66}\delta_n}$ it is enough to note that $\frac{\sum_{\ell<j^*}np_{n,\ell}}{n+2\sum_{\ell<j^*}np_{n,\ell}+n^{0.66}\delta_n}$ is an increasing function of $\sum_{\ell<j^*}np_{n,\ell}$ and therefore is minimized if and only if $\sum_{\ell<j^*}np_{n,\ell}$ takes its minimum value. However, according to Proposition 1 the minimum value of this term is $\Theta(n)$. Therefore, we have

$$\frac{\sum_{\ell<j^*}np_{n,\ell}}{n+2\sum_{\ell<j^*}np_{n,\ell}+n^{.66}\delta_n} - P(A^c)$$

$$\ge \frac{np_0}{np_0+n+n^{.66}\delta_n} - p(A^c) = \frac{p_0}{p_0+1}(1+o(1)).$$

27

This completes the proof. □

5. Tapered NLM Weights

In this section we show that the upper bound we provided for the NLM in Theorem 4 holds in the more general setting of tapered weights. We now allow the weights to be a smooth function of the δ_n-neighborhood. We assume that the weight assignment policy satisfies the following properties:

B1: The neighborhood size $\delta_n = 2\log(n)$.

B2: The weights are non-negative and bounded, i.e., $0 \leq w_{m,\ell} \leq \alpha$.

B3: If $d^2(x_{i,j}, x_{m,\ell}) = 0$, then the assigned weight satisfies $\mathbb{E}(w_{i,j}(m,\ell)) > c$, for some constant c independent of n.

B4: If $d^2(x_{i,j}, x_{m,\ell}) = 1$, then $\mathbb{E}(w_{i,j}(m,\ell)) = O\left(\frac{1}{\sqrt{n}}\right)$. It shall be empha-sized that slower decay rate in this expectation, results in slower decay in the rate of the NLM algorithm.

Theorem 7. *If the weight assignment policy in NLM satisfies properties B1–B4, then*

$$\sup_{f \in H^\alpha(C)} R(f, \hat{f}^N) = O\left(\frac{\log n}{n}\right).$$

Proof. Consider the four partitions S_1–S_4 defined in the proof of Theorem 4. Our goal is to obtain an upper bound for the risk of the pixels in each region. The risk of the pixels in S_2 and S_3 will be bounded by the strategy we employed in Theorem 4. Here, we just explain how we bound the risk of the pixels in S_1 and S_4. Since the proof for S_4 is the same as the proof for S_1, we consider just S_1. Let $(i,j) \in S_1$. Therefore $x_{i,j} = 0$ and

$$
\begin{aligned}
\mathbb{E}(x_{i,j} - \hat{f}_{i,j}^N)^2 &= \mathbb{E}\left(\frac{\sum w_{m,\ell} y_{m,\ell}}{\sum w_{m,\ell}}\right)^2 \\
&= \mathbb{E}\left(\frac{\sum w_{m,\ell} x_{m,\ell}}{\sum w_{m,\ell}}\right)^2 + \mathbb{E}\left(\frac{\sum w_{m,\ell} z_{m,\ell}}{\sum w_{m,\ell}}\right)^2 \\
&\quad + \mathbb{E}\left(\left(\frac{\sum w_{m,\ell} x_{m,\ell}}{\sum w_{m,\ell}}\right)\left(\frac{\sum w_{m,\ell} z_{m,\ell}}{\sum w_{m,\ell}}\right)\right). \quad (26)
\end{aligned}
$$

To obtain an upper bound for the risk, we will find upper bounds for the last three terms in (26). Lemmas 9 and 10 below summarize the upper bounds.

Lemma 9. *Let $w_{m,\ell}$ be the weights of NLM satisfying B1–B4. Then*

$$\mathbb{E}\left(\frac{\sum w_{m,\ell} x_{m,\ell}}{\sum w_{m,\ell}}\right)^2 = O\left(\frac{1}{n}\right).$$

Proof. Define S_f as the set of the indices of the pixels whose noise-free value is neither 0 nor 1, and plug in the actual values of $x_{m,\ell}$ to obtain

$$\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_1\cup S_4} w_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_2\cup S_3\setminus S_f} w_{m,\ell} x_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_1\cup S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_2\cup S_3\setminus S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell}}\right)^2$$

$$\leq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_1\cup S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_2\cup S_3} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell}}\right)^2$$

$$\leq \mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3} \alpha + 2n\alpha}{\sum_{(m,\ell)\in S_1\cup S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3} \alpha}\right)^2. \tag{27}$$

To derive the last inequality we use the following facts, which are easy to check:

1. $\left(\frac{\sum_{(m,\ell)\in S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_1\cup S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_2\cup S_3} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell}}\right)^2$ is an increasing function of $\sum_{(m,\ell)\in S_3\setminus S_f} w_{m,\ell}$.

2. $\left(\frac{\sum_{(m,\ell)\in S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell} x_{m,\ell}}{\sum_{(m,\ell)\in S_1\cup S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_2\cup S_3} w_{m,\ell} + \sum_{(m,\ell)\in S_f} w_{m,\ell}}\right)^2$ is a decreasing function of $\sum_{(m,\ell)\in S_2\setminus S_f} w_{m,\ell}$.

3. $|S_f| \leq 2n$, i.e., S_f contains at most $2n$ pixels.

Our next claim is that $\sum_{(m,\ell)\in S_4} w_{m,\ell}$ and $\sum_{(m,\ell)\in S_1} w_{m,\ell}$ concentrate around their means. We establish this in a manner very similar to the proof of Theorem 5. We first break the $\sum_{(m,\ell)\in S_4} w_{m,\ell}$ into $(4\delta_n + 2)^2$ subsequences such that each subsequece contains only independent random variables. In other words if $x_{m,\ell}$ is in one summation, then no other element of $\mathcal{C}_{i,j}^{4\delta_n+2}$ will be in the summation. Therefore, for each summation we can apply the Hoeffding inequality. Finally, we use the union bound as explained in the

proof of Theorem 5 to show that

$$\mathbb{P}\left(\left|\sum_{(m,\ell)\in S_1} w_{m,\ell} - \sum_{(m,\ell)\in S_1} \mathbb{E}(w_{m,\ell})\right| > t\right) \leq 2(4\delta_n + 2)^2 e^{\frac{-2t^2}{(4\delta_n+2)^4(\sum_{(m,\ell)\in S_1}\alpha^2)}},$$

$$\mathbb{P}\left(\left|\sum_{(m,\ell)\in S_4} w_{m,\ell} - \sum_{(m,\ell)\in S_1} \mathbb{E}(w_{m,\ell})\right| > t\right) \leq 2(4\delta_n + 2)^2 e^{\frac{-2t^2}{(4\delta_n+2)^4(\sum_{(m,\ell)\in S_1}\alpha^2)}}.$$

It is straightforward to prove that by setting t to $32\alpha n \log^{2.5}(n)$, we have

$$\mathbb{P}\left(\left|\sum_{(m,\ell)\in S_1} w_{m,\ell} - \sum_{(m,\ell)\in S_1} \mathbb{E}(w_{m,\ell})\right| > 32\alpha n \log^{2.5}(n)\right) \leq O\left(\frac{\delta_n^2}{n^8}\right)$$

$$\mathbb{P}\left(\left|\sum_{(m,\ell)\in S_4} w_{m,\ell} - \sum_{(m,\ell)\in S_1} \mathbb{E}(w_{m,\ell})\right| > 32\alpha n \log^{2.5}(n)\right) \leq O\left(\frac{\delta_n^2}{n^8}\right). \quad (28)$$

Define the event F as $\left\{\left|\sum_{(m,\ell)\in S_1} w_{m,\ell} - \sum_{(m,\ell)\in S_1} \mathbb{E}(w_{m,\ell})\right| < 32\alpha n \log^{2.5} n\right\}$
$\cap \left\{|\sum_{(m,\ell)\in S_4} w_{m,\ell} - \sum_{(m,\ell)\in S_1} \mathbb{E}(w_{m,\ell})| < 32\alpha n \log^{2.5} n\right\}$. It is clear from
(28) that

$$\mathbb{P}(\mathcal{F}^c) = O\left(\frac{\delta_n^2}{n^8}\right). \quad (29)$$

Using (27), (28), and (29) we have

$$\mathbb{E}\left(\frac{\sum_{(m,\ell)\in S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} \alpha + 2n\alpha}{\sum_{(m,\ell)\in S_1\cup S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} \alpha}\right)^2$$

$$\leq \mathbb{E}\left(\left(\frac{\sum_{(m,\ell)\in S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} \alpha + 2n\alpha}{\sum_{(m,\ell)\in S_1\cup S_4} w_{m,\ell} + \sum_{(m,\ell)\in S_3\setminus S_f} \alpha}\right)^2 \middle| F\right) \mathbb{P}(F) + \mathbb{P}(F^c)$$

$$\leq O\left(\frac{1}{n}\right).$$

The last inequality is due to Assumptions B3 and B4. This completes the
proof of the lemma. $\qquad\square$

30

Lemma 10. *Let $w_{m,\ell}$ be the weights of NLM with $\delta_n = \log(n)$. Also assume that the weights are set according to B1--B4. We then have*

$$\mathbb{E}\left(\frac{\sum w_{m,\ell} z_{m,\ell}}{\sum w_{m,\ell}}\right)^2 = O\left(\frac{\log^2(n)}{n^2}\right).$$

Proof. We first condition on the event F introduced in the proof of Lemma 9.

$$\mathbb{E}\left(\left(\frac{\sum w_{m,\ell} z_{m,\ell}}{\sum w_{m,\ell}}\right)^2\right)$$

$$\leq \mathbb{E}\left(\left(\frac{\sum w_{m,\ell} z_{m,\ell}}{\sum w_{m,\ell}}\right)^2 \,\bigg|\, F\right)\mathbb{P}(F) + \mathbb{P}(F^c)$$

$$\leq \mathbb{E}\left(\left(\frac{\sum w_{m,\ell} z_{m,\ell}}{\sum \mathbb{E}(w_{m,\ell}) - 32\alpha n \log^{2.5}(n)}\right)^2 \,\bigg|\, F\right)\mathbb{P}(F) + \mathbb{P}(F^c)$$

$$\leq \mathbb{E}\left(\left(\frac{\sum w_{m,\ell} z_{m,\ell}}{\sum \mathbb{E}(w_{m,\ell}) - 32\alpha n \log^{2.5}(n)}\right)^2\right) + \mathbb{P}(F^c)$$

$$\leq O\left(\frac{\log^2(n)}{n^2}\right).$$

The last inequality is due to the fact that

$$\mathbb{E}\left(\sum_{(m,\ell) \in S_{14}} w_{m,\ell} z_{m,\ell} \,\bigg|\, \mathcal{C}_{i,j}^{\delta_n}\right)^2$$

$$= \mathbb{E}\left(\sum_{(m,\ell) \in S_{14}} \sum_{(m',\ell') \in S_{14}} w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'} \,\bigg|\, \mathcal{C}_{i,j}^{\delta_n}\right)$$

$$= \sum_{(m,\ell) \in S_{14}} \sum_{(m',\ell') \in \mathcal{C}_{m,\ell}^{2\delta_n}} \mathbb{E}(w_{m,\ell} z_{m,\ell} w_{m',\ell'} z_{m',\ell'} \mid \mathcal{C}_{i,j}^{\delta_n}) = O(n^2 \delta_n^2).$$

Therefore,

$$\mathbb{E}\left(\left(\frac{\sum_{(m,\ell) \in S_{23}} w_{m,\ell} z_{m,\ell}}{\sum_{(m,\ell) \in S_{14}} w_{m,\ell}}\right)^2\right) \leq O\left(\frac{\delta_n^2}{n^2}\right). \tag{30}$$

$\square$

31

Using the bounds derived in Lemmas 9 and 10, we can complete the proof of the main theorem:

$$\sup_{f \in H^\alpha(C)} R(f, \hat{f}^{NL}) = \frac{1}{n^2} \sum_i \sum_j \mathbb{E}(x_{i,j} - \hat{f}_{i,j}^N)^2$$

$$\leq \frac{\log^2(n)(|S_1| + |S_4|)}{n^4} + \frac{|S_2| + |S_3|}{n^2} \leq O\left(\frac{\log(n)}{n}\right).$$

$\square$

6. Discussion

We have provided the first asymptotic result on the risk analysis of the nonlocal means (NLM) algorithm on smooth images with sharp edges. In contrast to most other filtering approaches, NLM does not consider the spatial vicinity of the pixels as a feature for setting the weights. Instead, it exploits more global features, which leads to improved performance.

In spite of this success, we have shown that the performance of NLM is within a logarithmic factor of the performance of the wavelet thresholding and still significantly below the optimal achievable rate. This is due to the fact that the isotropic nature of the NLM neighborhoods does not allow the algorithm to discriminate the pixels that are close to but below the edge from the pixels that are close to but above the edge. This leads to a blurring effect that results in high bias along the edge. Exploring the performance of NLM with anisotropic neighborhoods may address this issue and is left for future research.

7. Acknowledgements

References

[1] L. I. Rudin, S. Osher, E. Fatemi, Physica D: Nonlinear Phenomena 60 (1992) 259 – 268.

[2] D. L. Donoho, I. M. Johnstone, Annals of Statistics 26 (1998) 879–921.

[3] A. Korostelev, A. Tsybakov, Minimax theory of Image Reconstruction, Lecture Notes in Statistics, Springer-Verlag, 1993.

[4] D. L. Donoho, Annals of Statistics 27 (1999) 859 – 897.

[5] E. Arias-Castro, D. L. Donoho, Annals of Statistics 37 (2002) 1172–1206.

[6] L. Yaroslavsky, Digital image processing-An introduction, Springer Verlag, 1985.

[7] S. M. Smith, J. M. Brady, International Journal of Computer Vision 23 (1997) 45–78.

[8] C. Tomasi, R. Manduchi, in: International Conference on Computer Vision, pp. 839 –846.

[9] J.-S. Lee, Computer Vision, Graphics, and Image Processing 24 (1983) 255 – 269.

[10] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, CA, 1997.

[11] E. Candes, D. L. Donoho, Curvelets: A Surprisingly Effective Nonadaptive Representation of Objects with Edges, Technical Report, 1999.

[12] G. Kutyniok, D. Labatte, Journal on Wavelet Theory and Applications 1 (2007) 1–10.

[13] M. N. Do, M. Vetterli, IEEE Transactions on Image Processing 14 (2005) 2091–2106.

[14] A. Buades, B. Coll, J. Morel, SIAM Journal on Multiscale Modelling and Simulation 4 (2005) 490–530.

[15] J. M. Broder, New York Times (2011).

[16] C. Stein, Approximate computation of expectation, Institute of Mathematical Statistics, 1986.